Introduction to Data Science and Engineering

Project Proposal

Project Team

1. Ömer Faruk Satık – 150210330
2. Abdullah Aydoğan – 150230731
3. Selman Turan Toker – 150220330

Project Title

Relationship Analysis Between Electricity Generation, Greenhouse Gas Emissions and Deaths Due to Respiratory Diseases

Summary

The project aims to conduct an analysis examining possible relationships between these factors, using data on electrical energy production, greenhouse gas emissions and deaths due to respiratory disease. Three data sets will be used in the project. In the light of these data sets, first of all, the effect of the electrical energy production method on greenhouse gas emissions will be examined. Afterwards, the relationship between the change in greenhouse gas emissions over the years and the number of deaths due to respiratory diseases over the years will be studied. As a result, information will be obtained about the indirect effect of the energy production method on death. The project will be carried out according to the data obtained between 2009-2017. The reason for this is the lack of cumulative data sets on the internet regarding causes of death in Turkey. However, this situation is acceptable in this respect: Due to the global pandemic that emerged in 2019, respiratory-related deaths have increased significantly. For this reason, including the data of 2019 and subsequent years within the scope of the project may lead to data illusion.

Problem Definition

The main problem of the project is to understand the potential links between energy production and greenhouse gas emissions and to evaluate the effects of these factors on the number of deaths from respiratory disease. In this way, it will be possible to take the right steps in the energy production sector to improve the quality of life of citizens and even save their lives.

Datasets Sources

1. Electricity generation and shares by energy resources, 1970 - 2021

This data set shows the distribution of electrical energy production in Turkey and the energy sources used by years. The main energy sources included in the data set are:

- Coal: Shows the share of coal-based energy used in electrical energy production.
- Liquid Fuels: Shows the share of liquid fuels such as petroleum derivatives in electricity production.
- Natural Gas: Shows the share of natural gas in electrical energy production.
- Hydraulic (Hydro): It shows the share of hydraulic energy obtained using water power in electricity production.
- Renewable Energy and Wastes: Shows the total share of renewable energy sources such as geothermal, wind, solid biomass, solar, biogas and waste in electricity production.

The data set also provides total electrical energy production (in GWh). The shares of each energy source in total electricity production by year are expressed as percentages. This data set can be used to provide information about trends in energy consumption and use of energy resources in Turkey.

Source: TurkStat, TETC, Electricity Generation - Transmission Statistics of Turkey

2. Greenhouse gas emissions (CO2 equivalent), 1990 – 2021

This dataset is a table detailing greenhouse gas emissions (CO2 equivalent) by specific gas types by year. There are columns given below:

- Total: Total greenhouse gas emissions in a given year in million tons.
- CO2: Carbon dioxide emissions are in million tons.
- CH4: Methane emissions are in million tons.
- N2O: Emission of nitrogen oxides in million tons.
- F-gases: Emissions of greenhouse gases (F-gases) such as fluorinated gases are in millions of tons.

Source: TurkStat, Greenhouse Gas Emissions Statistics, 1990 – 2021

3. Distribution of selected causes of death by age group and gender, 2009-2017

This dataset represents a table containing death statistics by selected causes of death between 2009 and 2017. The data set provides detailed information across different age groups, genders and causes of death.

Source: TurkStat, Causes of Death Statistics

Possible Research Questions

1. What kind of impact does changing electrical energy production methods have on greenhouse gas emissions?
2. Is there a significant correlation between greenhouse gas emissions and deaths from respiratory diseases?
3. How does electricity generation based on different energy sources affect total respiratory disease mortality rates?
4. What are the effects of changes in energy policies during the 2009-2017 period on the factors considered?